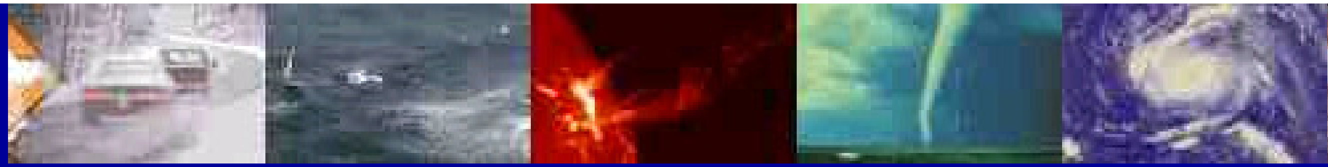


Decomposition and Attribution of Forecast Errors

Fanglin Yang

Environmental Modeling Center
National Centers for Environmental Prediction
College Park, Maryland, USA
Fanglin.yang@noaa.gov

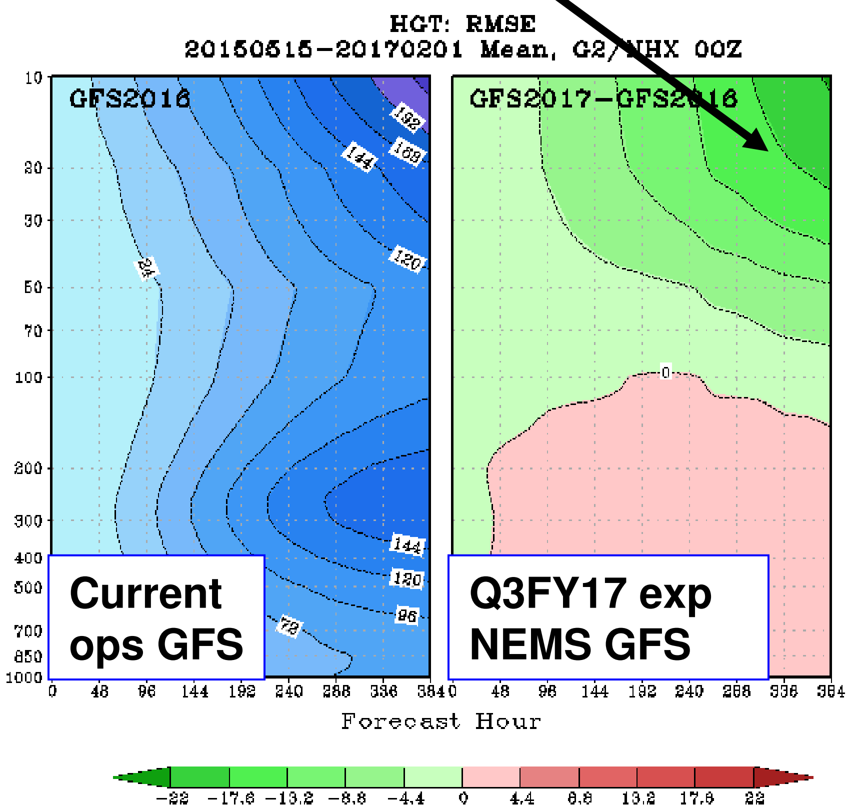
7th International Verification Methods Workshop
Berlin, Germany, May 3-11, 2017

Outline

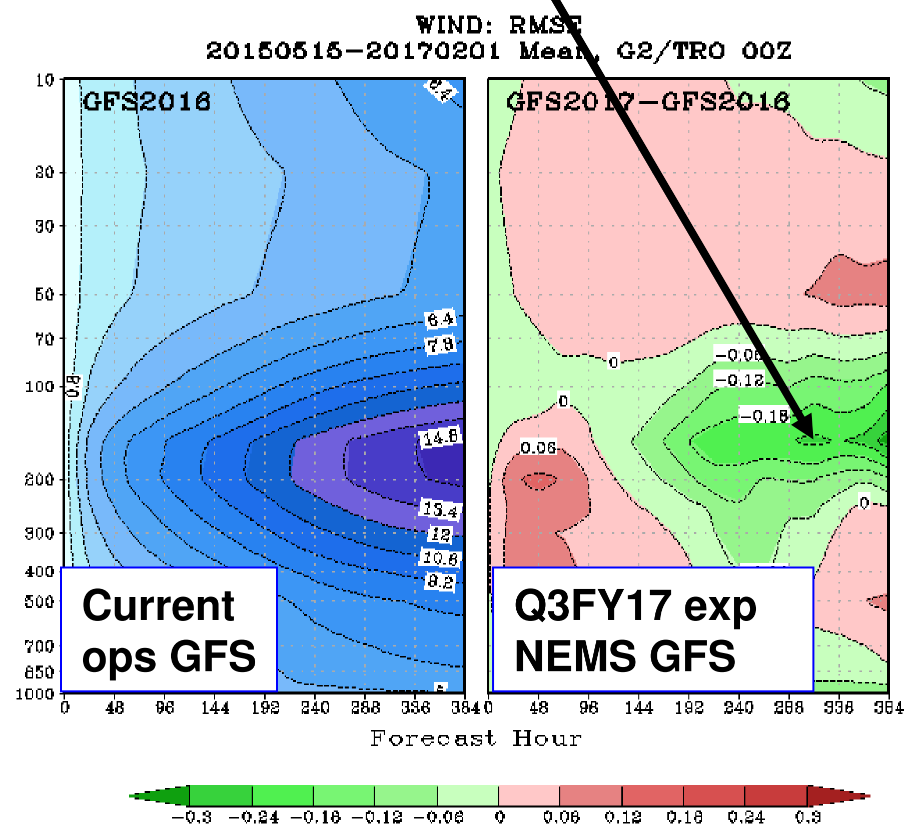
- Introduction to RMSE in model evaluation
- Decomposing RMSE for scalar variables
- Decomposing RMSE for vector winds
- A revised RMSE verification

RMSE Has long been used as a performance metric for model evaluation.

Smaller HGT RMSE, exp is better than ops

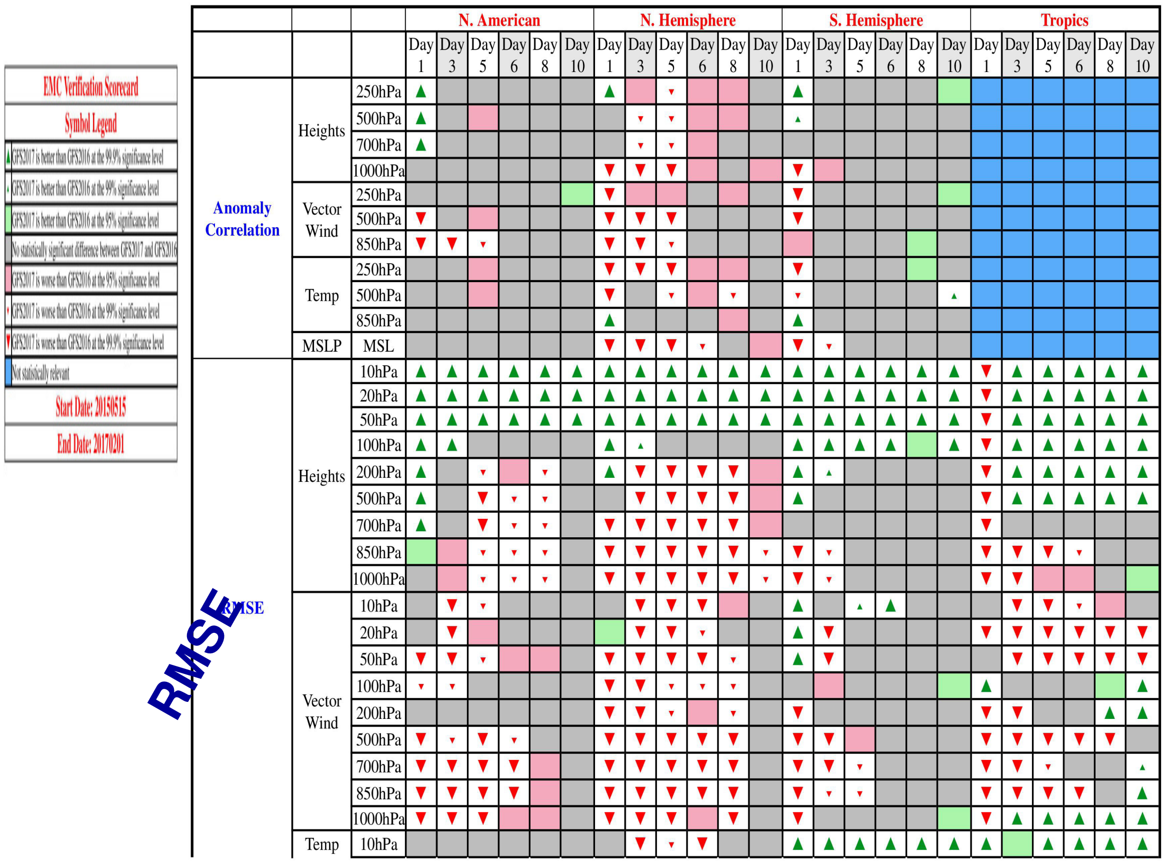


Smaller wind RMSE, exp is better than ops



NCEP-EMC GFS Verification Scorecard

ecard.html



1 of 3

5/6/17, 11:29 AM

Management often relies on the scorecard to make decision on model implementation

In this presentation I will demonstrate that RMSE can at times misrepresent model performance.

Root-Mean Squared Error (E) $$E = \sqrt{\frac{1}{n}\sum_{n=1}^{N}\left(F_n - A_n\right)^2}$$

Where, **F** is forecast, **A** is either analysis or observation, N is the total number of points in a temporal or spatial domain, or a spatial-temporal combined space.

$$\begin{aligned}
E^2 &= \frac{1}{n}\sum_{n=1}^{N}\left[\left(F_n - \overline{F}\right) - \left(A_n - \overline{A}\right) + \left(\overline{F} - \overline{A}\right)\right]^2 \\
&= \frac{1}{n}\sum_{n=1}^{N}\left(F_n - \overline{F}\right)^2 + \frac{1}{n}\sum_{n=1}^{N}\left(A_n - \overline{A}\right)^2 + \left(\overline{F} - \overline{A}\right)^2 \\
&\quad + \frac{2}{n}\cdot\sum_{n=1}^{N}\left[\left(F_n - \overline{F}\right) - \left(A_n - \overline{A}\right)\right]\cdot\left(\overline{F} - \overline{A}\right) \\
&\quad - \frac{2}{n}\cdot\sum_{n=1}^{N}\left(F_n - \overline{F}\right)\cdot\left(A_n - \overline{A}\right)
\end{aligned}$$

$$\boxed{E^2 = \sigma_f^2 + \sigma_a^2 - 2\sigma_f\sigma_a R + \left(\overline{F} - \overline{A}\right)^2}$$

Mean squared error

where

$$\sigma_f^2 = \frac{1}{n}\sum_{n=1}^{N}\left(F_n - \overline{F}\right)^2 \qquad \sigma_a^2 = \frac{1}{n}\sum_{n=1}^{N}\left(A_n - \overline{A}\right)^2$$

Variances of forecast & analysis

$$R = \frac{1}{n}\cdot\sum_{n=1}^{N}\left(F_n - \overline{F}\right)\cdot\left(A_n - \overline{A}\right) \Big/ \sigma_f\sigma_a$$

anomalous pattern correlation

$$E^2 = E_p^2 + E_m^2$$

Mean Squared Error: MSE

$$E_m^2 = \left(\overline{F} - \overline{A}\right)^2$$

MSE by Mean Difference

$$E_p^2 = \sigma_f^2 + \sigma_a^2 - 2\sigma_f \sigma_a R$$

MSE by Pattern Variation

Total MSE can be decomposed into two parts: the error due to differences in the mean and the error due to differences in pattern variation, which depends on standard deviation over the domain in question and anomalous pattern correlation to observation/analysis.

If a forecast has a larger mean bias than the other, its MSE can still be smaller if it has much smaller error in pattern variation, and vice versa.

In the following we discuss the characteristics of pattern variation

General Perception: **models with strong diffusion produce smoother fields, and hence have smaller RMSE.**
The answer is: **not always true**

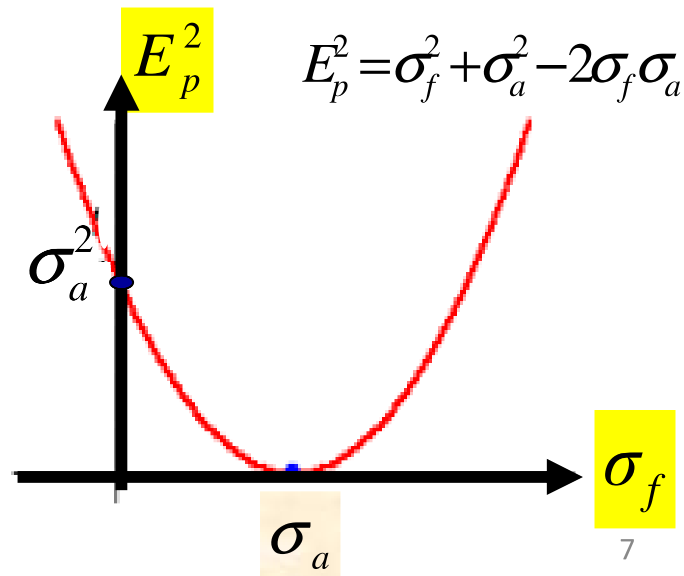
$$E_p^2 = \sigma_f^2 + \sigma_a^2 - 2\sigma_f \sigma_a R$$

$$\frac{\partial E_p^2}{\partial \sigma_f} = 2\sigma_f - 2\sigma_a R = 0 \qquad \Rightarrow E_p^2 \to \min \quad \text{if } \sigma_f = \sigma_a R$$

Case 1) R =1, perfect pattern correlation $E_p^2(\min) = 0 \quad \text{when } \sigma_f = \sigma_a$

One can see that if a forecast having either too large or too small a variance away from the analysis variance, its error of pattern variation increases.

$$R = 1 \qquad \Rightarrow \qquad E_p^2 = \left(\sigma_f - \sigma_a\right)^2$$

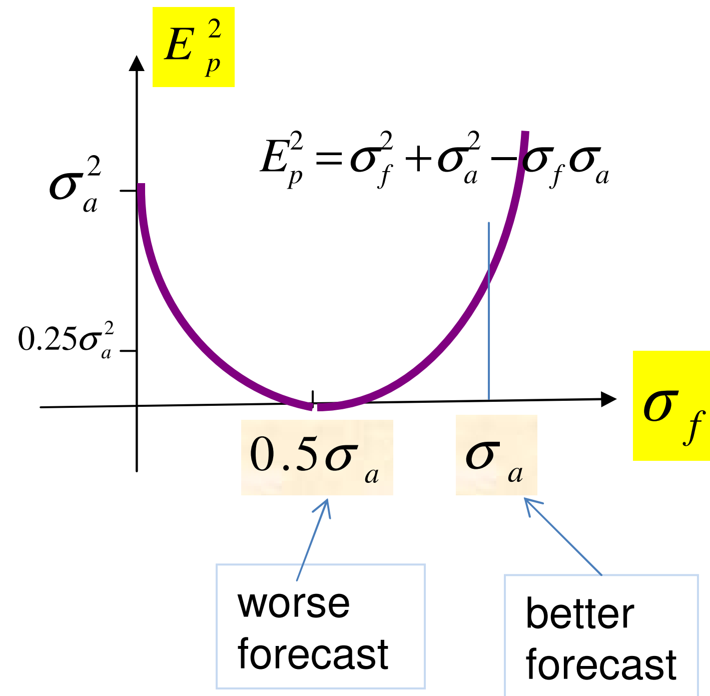


If R=1, E_p^2 does not award smooth forecasts that have smaller variances. It is not biased.

Case 2) R =0.5, imperfect pattern correlation

$$E_p^2(\min)=0 \quad \text{when } \sigma_f = 0.5\sigma_a$$

In this case, if one forecast has a better variance ($\sigma_f \to \sigma_a$) than the other ($\sigma_f \to 0.5\sigma_a$), the former will have a larger E_p^2 than the latter. **Good forecasts are actually penalized** !

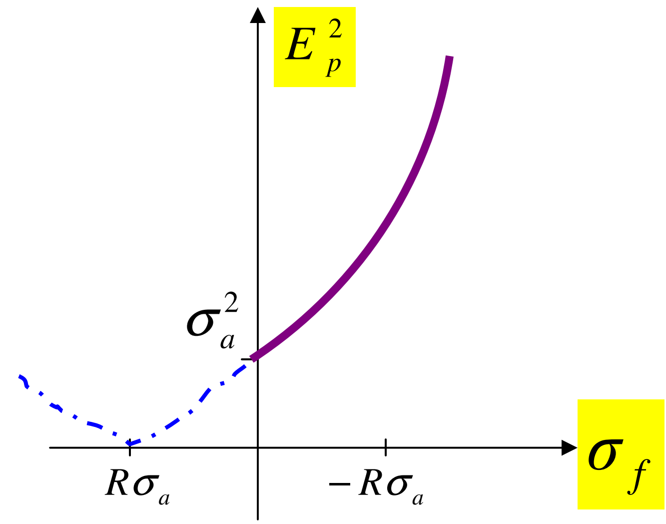


In general, if 0 < R < 1, E_p^2 awards smoother forecasts which have smaller variances close to $R\sigma_a$.

Case 3) For cases where $R \leq 0$,

$$E_p^2(\min) = \sigma_a^2 \quad \text{when} \quad \sigma_f = 0$$

E_p^2 **Increase monotonically with** σ_f



In this case, E_p^2 always awards smoother forecasts that have smaller variances .

Good forecast is penelized !

Will MSE normalized by analysis variance be unbiased?

$$E^2/\sigma_a^2 = E_p^2/\sigma_a^2 + E_m^2/\sigma_a^2$$

$$E_p^2/\sigma_a^2 = 1 - 2R\lambda + \lambda^2 \qquad \lambda = \sigma_f/\sigma_a$$

Assume

$$E_m^2 = 0$$

λ \ R	−1.0	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.2	1.44	1.36	1.28	1.20	1.12	1.04	0.96	0.88	0.80	0.72	0.64
0.4	1.96	1.80	1.64	1.48	1.32	1.16	1.00	0.84	0.68	0.52	0.36
0.6	2.56	2.32	2.08	1.84	1.60	1.36	1.12	0.88	0.64	0.40	0.16
0.8	3.24	2.92	2.60	2.28	1.96	1.64	1.32	1.00	0.68	0.36	0.04
1.0	**4.00**	**3.60**	**3.20**	**2.80**	**2.40**	**2.00**	**1.60**	**1.20**	**0.80**	**0.40**	**0.00**
1.2	4.84	4.36	3.88	3.40	2.92	2.44	1.96	1.48	1.00	0.52	0.04
1.4	5.76	5.20	4.64	4.08	3.52	2.96	2.40	1.84	1.28	0.72	0.16
1.6	6.76	6.12	5.48	4.84	4.20	3.56	2.92	2.28	1.64	1.00	0.36
1.8	7.84	7.12	6.40	5.68	4.96	4.24	3.52	2.80	2.08	1.36	0.64
2.0	9.00	8.20	7.40	6.60	5.80	5.00	4.20	3.40	2.60	1.80	1.00

$\lambda \to 1$

$E_p^2/\sigma_a^2 \downarrow$

R_c

$\lambda \to 1$

$E_p^2/\sigma_a^2 \downarrow$

- Ideally, for a given correlation R, the normalized error should always decrease as the ratio of forecast variance to analysis variance reaches to one from both sides. In the above table only when R is close to one (highly corrected patterns) does this feature exist. For most other cases, especially when R is negative, the normalized error decreases as the variance ratio decrease from two to zero. In other words, **the normalized error still favors smoother forecasts that have a variance smaller than the analysis variance (the truth).**

Is Mean-Squared-Error Skill Score (Murphy, MWR, 1988, p2419) Unbiased?

$$E^2/\sigma_a^2 = E_p^2/\sigma_a^2 + E_m^2/\sigma_a^2$$

$$E_p^2/\sigma_a^2 = 1 - 2R\lambda + \lambda^2 \qquad \lambda = \sigma_f/\sigma_a$$

$$MSESS = 1 - E^2/\sigma_a^2 = 2\lambda R - \lambda^2 - E_m^2/\sigma_a^2$$

Assume $E_m^2 = 0$

R

λ	– 1.0	– 0.8	– 0.6	– 0.4	– 0.2	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
0.2	– 0.44	– 0.36	– 0.28	– 0.20	– 0.12	– 0.04	0.04	0.12	0.20	0.28	**0.36**
0.4	– 0.96	– 0.80	– 0.64	– 0.48	– 0.32	– 0.16	0.00	0.16	0.32	0.48	**0.64**
0.6	– 1.56	– 1.32	– 1.08	– 0.84	– 0.60	– 0.36	– 0.12	0.12	0.36	0.60	**0.84**
0.8	– 2.24	– 1.92	– 1.60	– 1.28	– 0.96	– 0.64	– 0.32	0.00	0.32	0.64	**0.96**
1.0	**– 3.00**	**– 2.60**	**– 2.20**	**– 1.80**	**– 1.40**	**– 1.00**	**– 0.60**	**– 0.20**	**0.20**	**0.60**	**1.00**
1.2	– 3.84	– 3.36	– 2.88	– 2.40	– 1.92	– 1.44	– 0.96	– 0.48	0.00	–0.48	**0.96**
1.4	– 4.76	– 4.20	– 3.64	– 3.08	– 2.52	– 1.96	– 1.40	– 0.84	– 0.28	0.28	**0.84**
1.6	– 5.76	– 5.12	– 4.48	– 3.84	– 3.20	– 2.56	– 1.92	– 1.28	– 0.64	0.00	**0.64**
1.8	– 6.84	– 6.12	– 5.40	– 4.68	– 3.96	– 3.24	– 2.52	– 1.80	– 1.08	– 0.36	**0.36**
2.0	– 8.00	– 7.20	– 6.40	– 5.60	– 4.80	– 4.00	– 3.20	– 2.40	– 1.60	– 0.80	**0.00**

$\lambda \rightarrow 1$

$MSESS \uparrow$

R_c

$\lambda \rightarrow 1$

$MSESS \uparrow$

➢The best case is MSESS=1 when R=1 and Lambda=1. For most cases, especially when R is negative, MSESS decreases monotonically with Lambda. Therefore, MSESS **still favors smoother forecasts that have a variance smaller than the analysis variance.**

Summary I

- Conventional RMSE can be decomposed into *Error of Mean Difference* (Em) and *Error of Patter Variation* (Ep)

- Ep is unbiased and can be used as an objective measure of model performance *only if* the anomalous pattern correlation R between forecast and analysis is one (or very close to one)

- If R <1, Ep is biased and favors smoother forecasts that have smaller variances.

- Ep normalized by analysis variance is still biased and favors forecasts with smaller variance if anomalous pattern correlation is not perfect.

A complete model verification should include Anomalous Pattern Correlation, Ratio of Forecast Variance to Analysis Variance, Error of Mean Difference, and Error of Pattern Variation. RMSE can at times be misleading, especially when the anomalous pattern correlation between forecast and analysis is smaller.

$$E^2 = E_p^2 + E_m^2$$

$$E_p^2 = \sigma_f^2 + \sigma_a^2 - 2\sigma_f \sigma_a R$$

$$E_m^2 = \left(\overline{F} - \overline{A}\right)^2$$

Decomposing RMSE of Vector Wind

Vector Wind Stats

So far the deviations are for scalar variables. For vector wind, the corresponding stats are defined in the following way.

Define $\vec{V}_f = u_f \hat{i} + v_f \hat{j}$ $\quad \vec{V}_a = u_a \hat{i} + v_a \hat{j}$

Then **MSE:**

$$E^2 = \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \vec{V}_{an}\right)^2 = \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \vec{V}_{an}\right)\bullet\left(\vec{V}_{fn} - \vec{V}_{an}\right)$$

$$= \frac{1}{n}\sum_{n=1}^{N}\left(u_{fn}^2 + v_{fn}^2\right) + \frac{1}{n}\sum_{n=1}^{N}\left(u_{an}^2 + v_{an}^2\right) - \frac{2}{n}\sum_{n=1}^{N}\left(u_{fn}u_{an} + v_{fn}v_{an}\right)$$

$$= \mathrm{A} + \mathrm{B} - 2\mathrm{C}$$

where $A = \frac{1}{n}\sum_{n=1}^{N}\left(u_{fn}^2 + v_{fn}^2\right)$ $\quad B = \frac{1}{n}\sum_{n=1}^{N}\left(u_{an}^2 + v_{an}^2\right)$ $\quad C = \frac{1}{n}\sum_{n=1}^{N}\left(u_{fn}u_{an} + v_{fn}v_{an}\right)$

A, B, and C are partial sums in NCEP EMC VSDB database

Anomalous Pattern Correlation:

$$R = \frac{\frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \bar{\vec{V}}_{fn}\right)\bullet\left(\vec{V}_{an} - \bar{\vec{V}}_{an}\right)}{\sqrt{\frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \bar{\vec{V}}_{fn}\right)^2 \cdot \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{an} - \bar{\vec{V}}_{an}\right)^2}} = \frac{\sum_{n=1}^{N}\left[\left(u_{fn} - \bar{u}_f\right)\left(u_{an} - \bar{u}_a\right) + \left(v_{fn} - \bar{v}_f\right)\left(v_{an} - \bar{v}_a\right)\right]}{\sqrt{\sum_{n=1}^{N}\left[\left(u_{fn} - \bar{u}_f\right)^2 + \left(v_{fn} - \bar{v}_f\right)^2\right] \cdot \sum_{n=1}^{N}\left[\left(u_{an} - \bar{u}_a\right)^2 + \left(v_{an} - \bar{v}_a\right)^2\right]}}$$

Vector Wind Stats

$$E^2 = \frac{1}{n}\sum_{n=1}^{N}\left[\left(\vec{V}_{fn} - \overline{\vec{V}}_f\right) - \left(\vec{V}_{an} - \overline{\vec{V}}_a\right) + \left(\overline{\vec{V}}_f - \overline{\vec{V}}_a\right)\right]^2$$

$$= \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \overline{\vec{V}}_f\right)^2 + \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{an} - \overline{\vec{V}}_a\right)^2 + \left(\overline{\vec{V}}_f - \overline{\vec{V}}_a\right)^2$$

$$- \frac{2}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \overline{\vec{V}}_f\right)\bullet\left(\vec{V}_{an} - \overline{\vec{V}}_a\right)$$

$$\cancel{+ 2\left(\overline{\vec{V}}_f - \overline{\vec{V}}_a\right)\bullet\left\{\frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \overline{\vec{V}}_f\right) - \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{an} - \overline{\vec{V}}_a\right)\right\}}$$

$$= \sigma_f^2 + \sigma_a^2 - 2\sigma_f\sigma_a R + \left(\overline{u}_f - \overline{u}_a\right)^2 + \left(\overline{v}_f - \overline{v}_a\right)^2$$

$$= E_p^2 + E_m^2$$

where

$$E_m^2 = \left(\overline{\vec{V}}_f - \overline{\vec{V}}_a\right)^2 = \left(\overline{u}_f - \overline{u}_a\right)^2 + \left(\overline{v}_f - \overline{v}_a\right)^2$$

MSE by Mean Difference

$$E_p^2 = \sigma_f^2 + \sigma_a^2 - 2\sigma_f\sigma_a R$$

MSE by Pattern Variation

$$\sigma_f^2 = \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{fn} - \overline{\vec{V}}_f\right)^2 = \frac{1}{n}\sum_{n=1}^{N}\left[\left(u_{fn} - \overline{u}_f\right)^2 + \left(v_{fn} - \overline{v}_f\right)^2\right]$$

Variance of forecast

$$\sigma_a^2 = \frac{1}{n}\sum_{n=1}^{N}\left(\vec{V}_{an} - \overline{\vec{V}}_a\right)^2 = \frac{1}{n}\sum_{n=1}^{N}\left[\left(u_{an} - \overline{u}_a\right)^2 + \left(v_{an} - \overline{v}_a\right)^2\right]$$

Variance of analysis

Demonstration

Decomposed RMSE of Scalar and Vector Variables

Application to A Complete Objective Model Evaluation

Decomposing MSE of Scalar Variables

The following five components will be examined. All forecasts are verified against the same analysis, i.e., the mean of the two experiments pru12r and pre13d.

$$E^2 = E_p^2 + E_m^2$$

Total MSE

$$E_m^2 = \left(\overline{F} - \overline{A}\right)^2$$

MSE by Mean Difference

$$E_p^2 = \sigma_f^2 + \sigma_a^2 - 2\sigma_f \sigma_a R$$

MSE by Pattern Variation

$$\lambda = \frac{\sigma_f}{\sigma_a}$$

Ratio of Standard Deviation: Fcst/Anal

$$MSESS = 1 - E^2 / \sigma_a^2 = 2\lambda R - \lambda^2 - E_m^2 / \sigma_a^2$$

Murphy's Mean-Squared Error Skill Score

$$R = \frac{1}{n} \cdot \sum_{n=1}^{N} \left(F_n - \overline{F}\right) \cdot \left(A_n - \overline{A}\right) \Big/ \sigma_f \sigma_a$$

Anomalous Pattern Correlation

$$\sigma_f^2 = \frac{1}{n} \sum_{n=1}^{N} \left(F_n - \overline{F}\right)^2 \qquad \sigma_a^2 = \frac{1}{n} \sum_{n=1}^{N} \left(A_n - \overline{A}\right)^2$$

Decomposing RMSE of Vector Wind

The following five components will be examined. All forecasts are verified against the same analysis, i.e., the mean of the two experiments pru12r and pre13d.

$$E^2 = E_p^2 + E_m^2$$

Total MSE

$$E_m^2 = \left(\overline{u}_f - \overline{u}_a\right)^2 + \left(\overline{v}_f - \overline{v}_a\right)^2$$

MSE by Mean Difference

$$E_p^2 = \sigma_f^2 + \sigma_a^2 - 2\sigma_f \sigma_a R$$

MSE by Pattern Variation

$$\lambda = \frac{\sigma_f}{\sigma_a}$$

Ratio of Standard Deviation: Fcst/Anal

$$MSESS = 1 - E^2/\sigma_a^2 = 2\lambda R - \lambda^2 - E_m^2/\sigma_a^2$$

Murphy's Mean-Squared Error Skill Score

$$R = \frac{\frac{1}{n}\sum_{n=1}^{N}\left[\left(u_{fn} - \overline{u}_f\right)\left(u_{an} - \overline{u}_a\right) + \left(v_{fn} - \overline{v}_f\right)\left(v_{an} - \overline{v}_a\right)\right]}{\sigma_f \cdot \sigma_a}$$

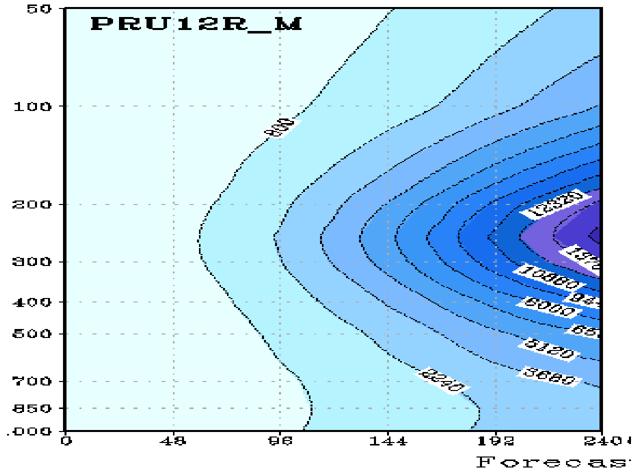
Anomalous Pattern Correlation

$$\sigma_f^2 = \frac{1}{n}\sum_{n=1}^{N}\left[\left(u_{fn} - \overline{u}_f\right)^2 + \left(v_{fn} - \overline{v}_f\right)^2\right]$$

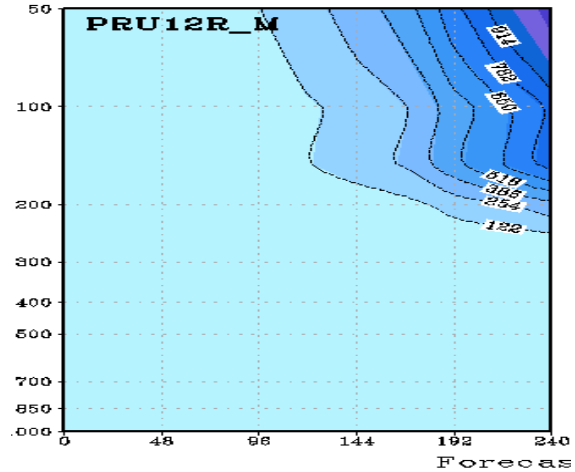
$$\sigma_a^2 = \frac{1}{n}\sum_{n=1}^{N}\left[\left(u_{an} - \overline{u}_a\right)^2 + \left(v_{an} - \overline{v}_a\right)^2\right]$$

Decomposing NH HGT MSE, T382L64 GFS

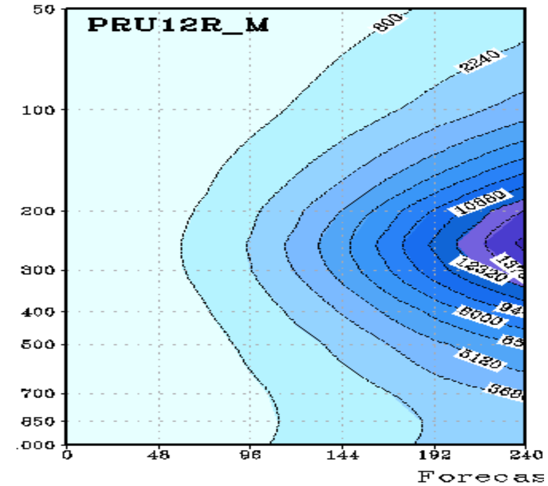
Total MSE



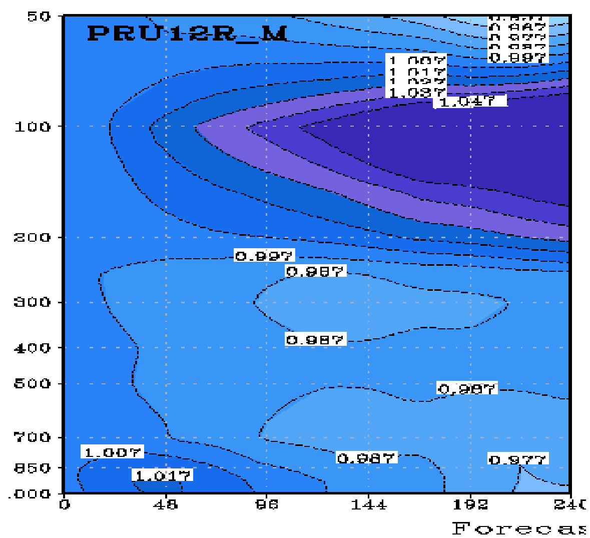
MSE by Mean Difference



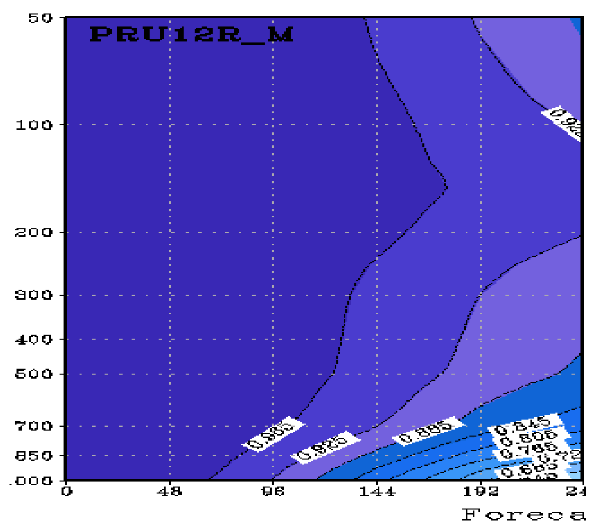
MSE by Pattern Variation



Ratio of Standard Deviation



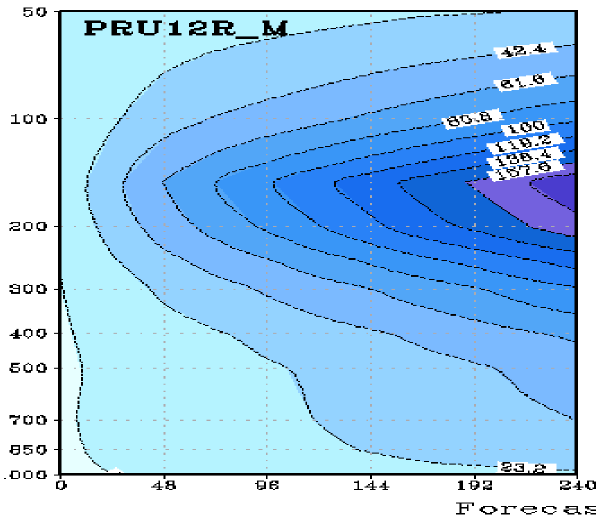
Anomalous Pattern Correlation



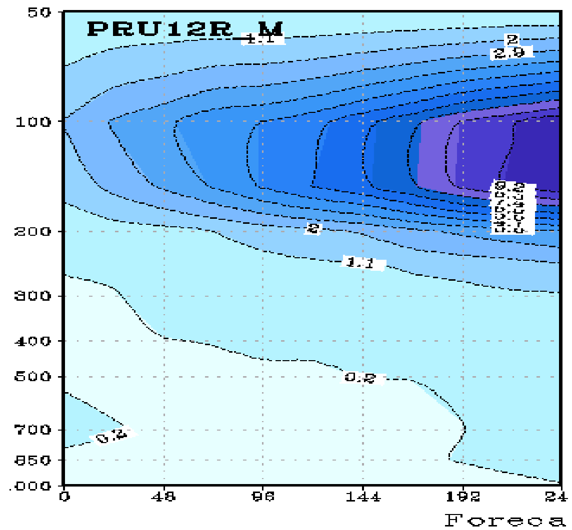
- Total RMSE is primarily composed of EMD in the lower stratosphere and EPV in the troposphere.
- HGT generally has high anomalous pattern correlation.
- The forecast variance is lower than that of analysis in the lower troposphere and stratosphere, and larger near the tropopause.
- Forecast variance near tropopause increases with forecast lead time .

Decomposing Tropical Vector Wind RMSE^2, T382L64 GFS

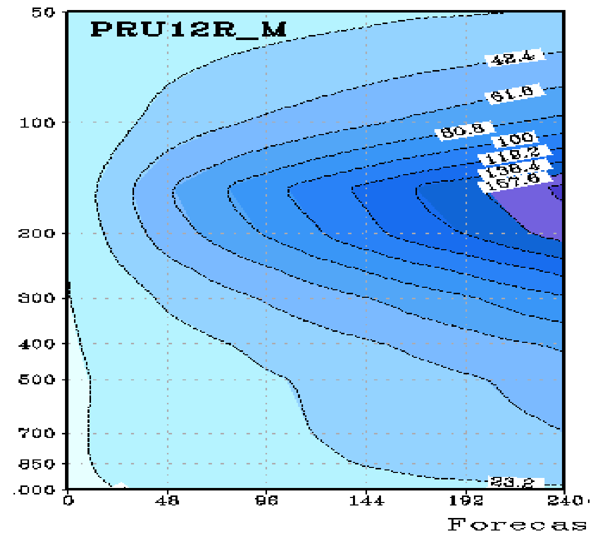
Total MSE



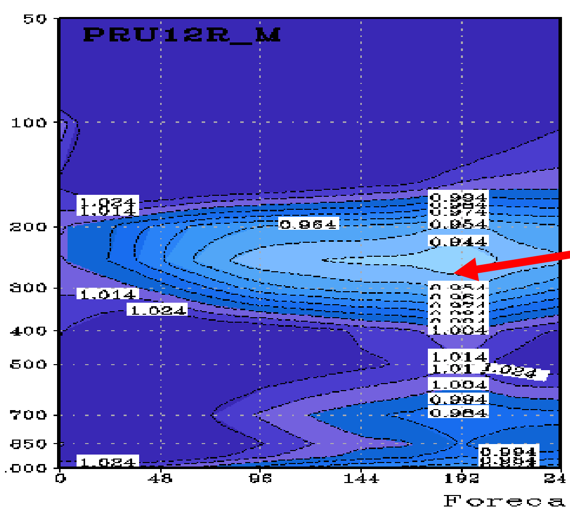
MSE by Mean Difference



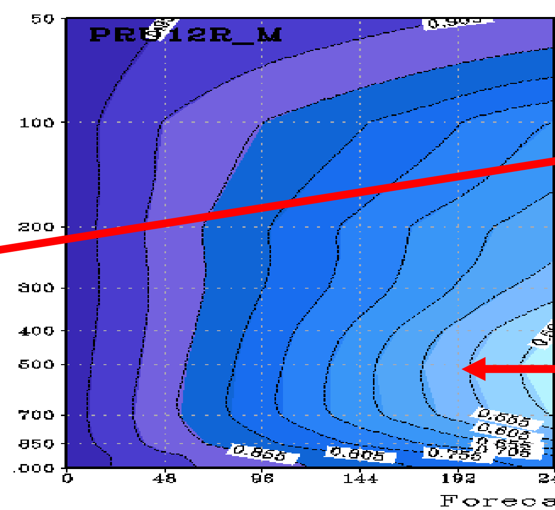
MSE by Pattern Variation



Ratio of Standard Deviation



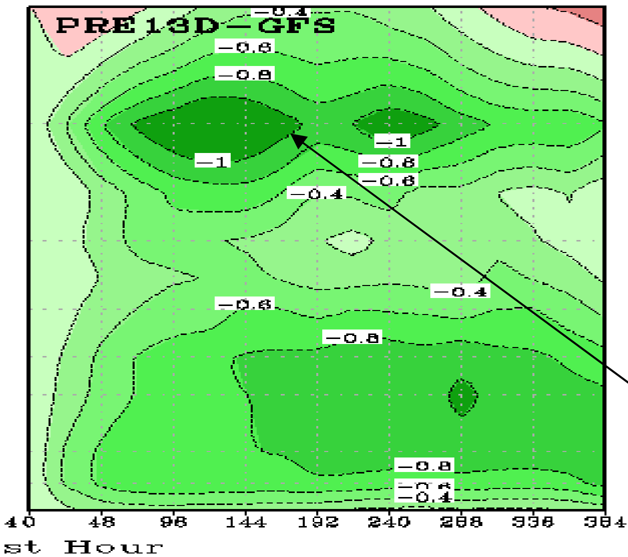
Anomalous Pattern Correlation



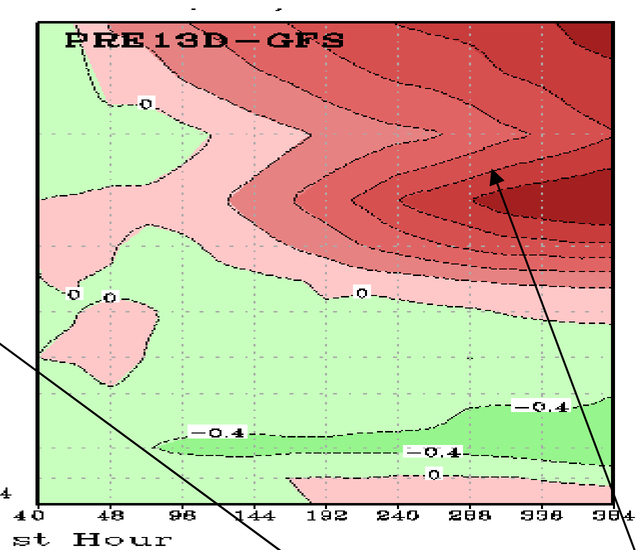
- **For tropical Wind, both EMD and EPV are concentrated near the tropopause , and increase with forecast lead time.**
- **T382 GFS is not able to maintain wind variance near the tropopause, and has stronger variance everywhere else.**
- **Wind anomalous pattern correlation is much poorer than that of HGT, and faints quickly with forecast lead time, especially in the lower troposphere.**

Tropical Vector Wind RMSE, T574GFS - T382GFS, Q3FY2010 Implementation

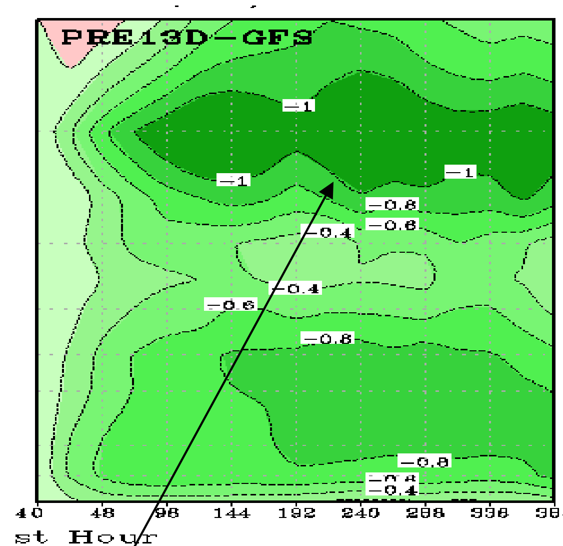
Total MSE



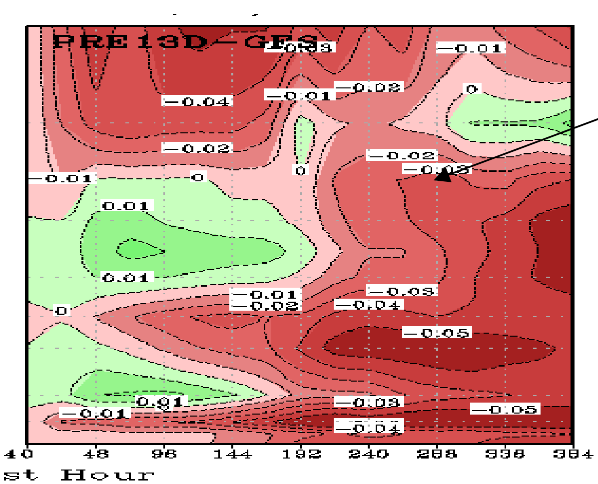
MSE by Mean Difference



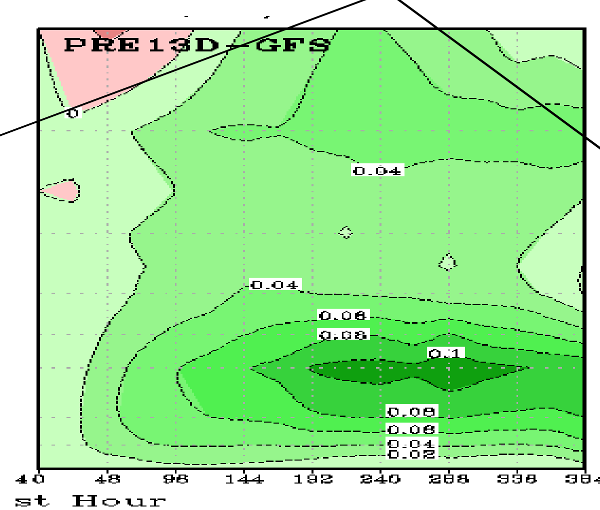
MSE by Pattern Variation



Ratio of Standard Deviation



Anomalous Pattern Correlation



T475 has weaker winds due to stronger background diffusion.

RMSE from pattern variation is reduced because wind is smoother. Thus overall smaller RMSE (misleading ?)

Summary

- **RMSE/MSE can be at times misleading. Its fairness as a performance metric depends on the goodness of mean difference, standard deviation, and pattern correlation.**

- **If pattern correlation is low, RMSE tends to award forecasts with smoother fields. The implication is that RMSE should not be used for extended NWP forecasts and seasonal forecasts either.**

- **RMSE has often been used as the only metric to measure model forecast performance in the tropics, especially for wind forecast. A more comprehensive verification should at least include MSE, MSE by Mean Difference, Anomalous Pattern Correlation, and Ratio of Forecast Variance to Analysis Variance.**